Integrated Intelligent Research (IIR)
International Journal of Data Mining Techniques and Applications
Volume: 05 Issue: 02 December 2016 Page No.159-166
ISSN: 2278-2419

Finding Influential Healthcare Interventions of Different Socio-Economically and Educationally Segmented Regions by using Data Mining Techniques: Case Study on Nine High Focus States of India

Partha Saha, Uttam Kumar Banerjee
Research Scholar, Ranbir and Chitra Gupta School of Infrastructure Design and Management Indian Institute of Technology, Kharagpur, India
Email: partha.saha@iitkgp.ac.in

Abstract - United Nations at Millennium Summit 2000 made targets on Under-five Mortality Ratio (U5MR) and Maternal Mortality Ratio (MMR) for improving health condition of mothers and children. Though India did not be able to achieve those targets but have improved significantly. Aim of the study is to find out influential healthcare interventions of socio-economically and educationally different regions which have high impact on their HIs. At resource constrained condition, strategic evidence based planning will help healthcare department to reduce inequity in HIs among different regions. Data of different HIs has been collected from Family Welfare Statistics of India 2012 and healthcare interventions have been collected from District Level Household Survey 3. 192 districts from 'Nine High Focus States of India' have been used as case study area in this research work. Both hierarchical and k-means, clustering techniques have been used for segmenting 192 districts based on their socio-economic and educational status and decision tree classification technique has been used for building relationship model for each segment. Total six decision tree classifiers have been developed for identifying most influential interventions on Infant Mortality Rate (IMR) and U5MR. From this work it has become clear that impact of healthcare interventions on healthcare indicators varies from region to region. In hilly regions, adolescent interventions had more impact on U5MR and IMR than child age interventions.

Keywords - Maternal and Child Healthcare, Healthcare Interventions, Clustering technique, Decision Tree technique, High focus states of India

I. INTRODUCTION

In presence of officials of 189 countries at the Millennium Summit of United Nations (UN), eight targets were set up for eradicating extreme poverty, upholding human dignity, and abolishing inequity in basic human rights such as health, education etc. and those targets were titled as Millennium Development Goals (MDGs). Among eight MDGs, fourth and fifth MDGs were reduction of child mortality and improvement of maternal health respectively. Fourth MDG was reduction of under-five mortality by two thirds in between 1990 and 2015. Fifth MDG was divided into two sub-targets. Those were reduction of maternal mortality by three quarters in between 1990 and 2015 and achievement of universal access to reproductive health by 2015[1].

In India, U5MR has decreased from 114.3 (112.0-116.7) per 1000 live births on 1990 to 62.6 (58.2-67.3) per 1000 live births on 2010[2]. MMR has also improved from 523 (310-835) per 100000 live births on 1990 to 254 (154-395) per 100000 live births on 2008[3]. Still huge regional differences exist within India. Prominent difference can be seen between northern and southern part of India. U5MR in some of the states of northern India like Uttar Pradesh (94 per 1000 live births), Madhya Pradesh (89 per 1000 live births), Orissa (82 per 1000 live births), Assam and Bihar (77 and 78 per 1000 live births) were almost similar to the U5MR in some African countries[4]. Due to unacceptably high mortality rate, eight empowered action group states (Bihar, Chhattisgarh, Jharkhand, Madhya Pradesh, Orissa, Rajasthan, Uttrakhand and Uttar Pradesh) and Assam have been designated as 'High Focus states' by the Government of India[5]. Around 45% of India's geographical area and 48% of India's population stay in those states[6].

Inequity in coverage of maternal and child healthcare services is a major cause of poor healthcare indicators at high focus states of India. Determinants identified by several studies for inequity among healthcare services were education and socio-economic status of people. To improve maternal and child healthcare condition throughout every places, region-specific interventions should be designed and implemented. Among high focus states also, region-specific healthcare planning are required. Realising lack of research into this particular problem, this work has been performed. Aim of the study is to find out influential interventions of socio-economically and educationally different regions within high focus states of India which have high impact on IMR and U5MR of those regions. First 192 districts of nine high focus states of India have been clustered into three segments based on their socio-economic and educational condition by applying hierarchical clustering technique and k-means clustering technique. Then decision tree classification technique has been used to find out most influential interventions on healthcare indicators from each segment. Decision tree model has been developed for both IMR and U5MR and for all segments separately. Each model has been represented through decision tree diagram which helped to find out most influential interventions for each region for predicting mortality rates. These region specific knowledges will help policy makers to take strategic actions for improvement of child health system condition.

II. DATA & METHODS

2.1 Data

Data used in this study were all secondary data. Three broad categories of data - maternal and child healthcare indicators, socio-economic and educational parameters, and maternal and

International Journal of Data Mining Techniques and Applications
Volume: 05 Issue: 02 December 2016 Page No.159-166
ISSN: 2278-2419

child healthcare interventions have been used for analysis. Healthcare Indicators (HIs) - IMR, and U5MR of 2007-09 of 192 districts of nine high focus states of India have been collected from Family Welfare Statistics in India 2011[7]. For educational scenario analysis, variables used were male literacy rate, female literacy rate, percentage of boys attending school, and percentage of girls attending school. Availability of basic physical infrastructure like electricity, water, toilet, clean cooking fuel data have been scrutinized for getting knowledge about condition of each district and for economic condition, availability of assets have been analyzed. For understanding financial position, variables used in analysis were ownership of house, agricultural land, possession of television, mobile, and motorized vehicles.

Eminent epidemiologists like Barker, Wadsworth, and others discovered that diseases at adult age have relationship with events experienced at fetal age[7]. Researchers understood the requirement of healthcare interventions for all stages of life of a female, starting from adolescent age to pregnancy to childbirth. To comprehend child healthcare condition of nine high focus states of India, data of total 26 different healthcare interventions have been studied. Among them, six interventions are used by adolescent and reproductive age group, six interventions are accessed by mothers at the time of pregnancy and immediately after delivery, six interventions are for newborns, four interventions are for children aged less than five years, and four interventions are for improving knowledge of HIV/AIDS and RTI among adolescent and reproductive aged persons. Details of the interventions along with healthcare indicators and socio -economic, and educational parameters have been listed in table 1. Data of socio-economic and educational parameters and healthcare interventions were obtained from 3rd round (2007-08) of District Level Household and Facility Survey (DLHS)[8]. 1st round of the survey was taken place in 1998-99 and then second, and third round of surveys were conducted on 2002-03, and 2007-08 respectively. These surveys were organized by Ministry of Health and Family Welfare, Government of India to gain insights about quality and coverage of healthcare services along with condition of healthcare delivery points at districts in India. DLHS3 was conducted in total 601 districts of 34 states in India. 720320 households and 643944 ever-married women aged 15-49 years have been surveyed through this nationally representative survey. A multi-stage stratified systematic sampling design technique was used for these surveys. In this study, data of rural regions have been selected for analysis. Short abbreviated forms of variables have been used in this paper for saving space. All variables with their abbreviated names have been presented in table no. 1.

Table 1.List of variables with their abbreviated forms

Abbreviation	Details of Variables
Healthcare Indicators	
IMR	Infant Mortality Ratio per 1000 live births
U5MR	Under-five Mortality Ratio per 1000 live births
Socio-Economic and Educational Variables (all data are in percentage)	
Mlit	Male Literacy (age 7+)
Flit	Female Literacy rate (age 7+)
Gschl	Girls attending school (age 6-11)
Bschl	Boys attending school (age 6-11)
Elec	Having electricity connection
Tlt	Have Access to toilet facility
Wtr	Improved source of drinking water
LPG	Use LPG for cooking
Hse	Live in a pucca house
Own	Own a house
Agri	Own Agriculture Land
TV	Have a television
Mbl	Have a mobile phone
Vhcl	Have a Motorized Vehicle
Adolescent and Reproductive Age Healthcare Interventions (Utilized by currently married women (age 15-49)(all data are in percentage)	
FP	Usage of any family planning method
Fstrl	Female Sterilization
IUD	Intrauterine Device
Pill	Contraceptive Pill
Cndm	Condom
UN	Unmet Need for Family Planning
Maternal healthcare Interventions before and after delivery (all data are in percentage)	
Rgstrd	Mothers registered in the first trimester when they were pregnant with last live birth/still birth
ANC3	Mothers who had at least 3 ante-natal care visits during the last pregnancy
TTI	Mothers who got at least one TT injection when they were pregnant with their last live birth / still birth
Inst.dlvry	Institutional births
SBA	Delivery at home assisted by a doctor/nurse/LHV/ANM
PNC.mthr	Mothers who received post-natal care within 48 hours of delivery of their last child
Newborn healthcare Interventions (all data are in percentage)	
Full.imun	Children (12-23 months) fully immunized (BCG, 3 doses each of DPT and Polio, and Measles)
BCG	Children (12-23 months) who have received BCG Vaccine
Polio	Children (12-23 months) who have received 3 doses of Polio Vaccine
DPT	Children (12-23 months) who have received 3 doses of DPT Vaccine
Measles	Children (12-23 months) who have received Measles Vaccine
Vit.A	Children (9-35 months) who have received at least one dose of Vitamin A
Children Healthcare Interventions(all data are in percentage)	
ORS	Children with Diarrhea in the last two weeks who received ORS
Drh.trmnt	Children with Diarrhea in the last two weeks who were given treatment
ARI.trmnt	Children with acute respiratory infection /fever in the last two weeks who were given treatment

International Journal of Data Mining Techniques and Applications
Volume: 05 Issue: 02 December 2016 Page No.159-166
ISSN: 2278-2419

PNC.chld	Children had check-up within 24 hours after delivery (based on last live birth)
Knowledge of HIV/AIDS (Utilized by currently married women (age 15-49)(all data are in percentage)	
HIV.hrd	Women heard of HIV/AIDS
HIV.cndm	Women who knew that consistent condom use can reduce the chances of getting HIV/AIDS
HIV.tst	Women ever underwent test for detecting HIV/ AIDS
RTI.hrd	Women heard of RTI/STI

2.2 Methods

Aim of the study is to find region wise influential MCH interventions which have greater impact on IMR and U5MR. Total work of this project has been divided into three consecutive objectives. First, all 192 districts of nine high focus states have been segmented based on their socio-economic and educational condition. For segmenting districts, both hierarchical clustering technique and k-means clustering technique both have been used. In second phase, comparative position of variables among segments has been visualized. Two histograms have been plotted for comprehending health indicators position and healthcare interventions condition of all segments separately. Then in third objective, decision tree classification technique has been used for developing a relationship model among healthcare indicators and healthcare interventions. Total six classifiers have been generated for both IMR, U5MR and for all three segments. Details of the techniques used in this work have been described in following subsections.

2.2.1 Hierarchical Clustering Technique

In hierarchical clustering method, data objects are grouped with each other hierarchically and develop a tree of clusters. This technique works either in agglomerative manner or in divisive manner. Agglomerative process is a bottom-up approach. Here all individual objects are considered as individual cluster and then through the process of merge operations, each cluster groups with other clusters to make a bigger one. This process of merging happens hierarchically and ends when entire dataset becomes one cluster. Divisive process follows top-down approach. Here, entire dataset is first considered as one big cluster and then gradually it gets divided into smaller clusters. The process of division continues until clusters split up to individual data objects. In hierarchical clustering technique, the process of formation of clusters is represented by a tree like structure, called dendrogram. This cluster formation procedure requires both distance measure and linkage criterion parameters. In this study, average distance technique has been acquired.

2.2.2 K-means clustering Technique

Through k-means clustering technique user can divide the entire data set into desired number of clusters. This process primarily seeks cluster number and since it can be any non-zero value k, this technique is called as k-means clustering technique. Process of this technique is as follows. First, k objects are selected randomly from dataset as cluster mean for k clusters. Most similar objects are assigned into any of the k clusters. Then again mean values of all k clusters are computed. Now based on new means, all data are again reshuffled among clusters. This process of computing mean values and reshuffling of data continues until there is no change between new and previous mean value and new and old segments.

2.2.3 Decision Tree Technique

Decision tree is one of the most popular algorithms among supervised classification data mining techniques which develops hierarchical relationship model among response variable and explanatory variables. Quinlan's ID3, C4.5, C5 algorithms[9] and Breiman's Classification and Regression Tree (CART)[10] are maximum used for developing decision trees. As name suggests, in this technique observations are separated into branches to enhance predictability quality. The separation at a particular variable is decided by using mathematical algorithms (e.g. information gain, Gini index, and chi-squared test). Threshold values for all variables are computed and then the split position is decided. By repeating this process, all variables of same classes are grouped together. The most commonly used mathematical algorithm for splitting includes entropy based information gain (used in ID3, C4.5, C5), Gini index (used in CART), and chi-squared test (used in CHAID). CART algorithm is used to develop decision tree in this research work.

Fully built decision tree comes up with many branches that may reduce accuracy of the model. Tree pruning is required for improving predictability of the classifier. Through pruning, branches of the fully built tree are cut up to the optimal size where prediction error of the model becomes lowest. There are many processes for pruning. In this study, cross-validation cost (or CV cost) minimization process has been chosen for finding the optimal tree diagram. If response variable is categorical by nature then decision tree is named as classification tree and for continuous variable it is known as regression tree. Since in this study data type of response variables is continuous type, from now onwards, decision trees in this study will be designated as regression trees. R statistical software version 3.1.1 has been used for developing regression tree diagrams. Prediction accuracy of models for each category has been computed in this research paper. Prediction accuracy of each model has been measured by calculating Mean Square Error (MSE) among original and predicted values of response variable.

III. RESULTS AND DISCUSSIONS

3.1 Segmentation of Districts

Hierarchical clustering technique has been used for clustering all districts based on their socio-economic and educational condition. List of variables used for computing hierarchical clustering technique have been shown in table 1 under socio-economic and educational variables category. Agglomerative technique has been used as clustering technique and euclidian method has been used for distance measurement among clusters in this study. Dendrogram developed through hierarchical clustering has been shown in figure 1. From figure it can be observed that all districts at height zero were independent entity and gradually as height increased, they clubbed with other districts to form clusters. At height 100, five distinct clusters can be observed. Above that, first cluster and second cluster from left have merged into one cluster and fourth and fifth clusters have also united to form another cluster. Only the mid cluster or third cluster from left with lesser number of districts made a separate cluster. Finally, first

International Journal of Data Mining Techniques and Applications
Volume: 05 Issue: 02 December 2016 Page No.159-166
ISSN: 2278-2419

and second clusters integrated to form one cluster and met with the last group.

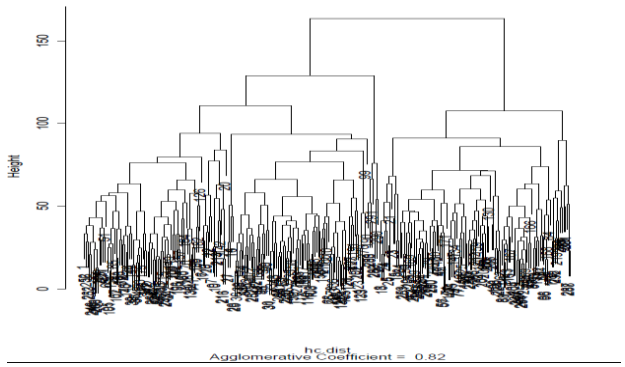

Figure 1: Dendrogram for hierarchical clustering of districts

Since district numbers are not visible from the above figure, k-means clustering technique has been used to identify districts under each segment. Same 14 socio-economic and educational variables have been used for segmentation analysis too. Understanding the nature in which districts can be grouped with each other, three clusters of districts have been developed by using k-means technique. List of districts under each segment has been shown in table 2 with their states.

Table 2: List of districts under three segments

Segment No.	State	Districts
Segment 1	Assam (5/27)	Kokrajhar, Nalbari, Lakhimpur, Hailakandi, Udalguri.
	Bihar (37/37)	PashchimChamparan, PurbaChamparan, Sheohar, Sitamarhi, Madhubani, Supaul, Araria, Kishanganj, Purnia, Katihar, Madhepura, Saharsa, Darbhanga, Muzaffarpur, Gopalganj, Siwan, Saran, Vaishali, Samastipur, Begusarai, Khagaria, Bhagalpur, Banka, Munger, Lakhisarai, Sheikhpura, Nalanda, Patna, Bhojpur, Buxar, Kaimur (Bhabua), Rohtas, Jehanabad, Aurangabad, Gaya, Nawada, Jamui
	Chhattisgarh (5/16)	Koriya, Surguja, Kanker, Bastar, Dantewada.
	Jharkhand (20/22)	Garhwa, Palamu, Hazaribagh, Kodarma, Giridih, Deoghar, Godda, Sahibganj, Pakaur, Dhanbad, Bokaro, Ranchi, Lohardaga, Gumla, PashchimiSinghbhum, PurbiSinghbhum, Simdega, Seraikela, Latehar, Jamtara.
	Madhya Pradesh (6/45)	Sheopur, Datia, Umaria, Shahdol, Narsimhapur, Balaghat.
	Odisha (18/30)	Jharsuguda, Sambalpur, Debagarh, Sundargarh, Kendujhar, Dhenkanal, Anugul, Ganjam, Gajapati, Kandhamal, Baudh, Sonapur, Balangir, Nuapada, Kalahandi, Rayagada, Nabarangapur, Koraput.
	Rajasthan (7/32)	Bharatpur, Dhaulpur, Karauli, Jaisalmer, Tonk, Rajsamand, Dungarpur.
	Uttar Pradesh (49/70)	Bijnor, Moradabad, Rampur, Gautam Buddha Nagar, Agra, Firozabad, Etah, Mainpuri, Budaun, Bareilly, Pilibhit, Shahjahanpur, Kheri, Sitapur, Hardoi, Unnao, Lucknow, Rae Bareli, Farrukhabad, Etawah, Auraiya, Kanpur Dehat, Jhansi, Lalitpur, Hamirpur, Mahoba, Banda, Chitrakoot, Fatehpur, Pratapgarh, Kaushambi, Allahabad, Barabanki, Faizabad, Ambedaker Nagar, Sultanpur, Bahraich, Shrawasti, Balrampur, Gonda, Siddharthnagar, Basti, SantKabir Nagar, Maharajganj, Gorakhpur, Kushinagar, Mau, Jaunpur, Mirzapur.
Segment 2	Chhattisgarh (11/16)	Jashpur, Raigarh, Korba, Janjgir-Champa, Bilaspur, Kawardha, Rajnandgaon, Durg, Raipur, Mahasamund, Dhamtari.
	Jharkhand (2/22)	Chatra, Dumka.
	Madhya Pradesh (39/45)	Morena, Bhind, Gwalior, Shivpuri, Guna, Tikamgarh, Chhatarpur, Panna, Sagar, Damoh, Satna, Rewa, Sidhi, Neemuch, Mandsaur, Ratlam, Ujjain, Shajapur, Dewas, Jhabua, Dhar, Indore, West Nimar, Barwani, East Nimar, Rajgarh, Vidisha, Bhopal, Sehore, Raisen, Betul, Harda, Hoshangabad, Katni, Jabalpur, Dindori, Mandla, Chhindwara, Seoni.
	Odisha (10/32)	Bargarh, Baleshwar, Bhadrak, Kendrapara, Jagatsinghapur, Cuttack, Jajapur, Nayagarh, Khordha, Puri.
	Rajasthan (22/32)	Hamumangarh, Bikaner, Jhunjhunun, Alwar, SawaiMadhopur, Dausa, Sikar, Nagaur, Jodhpur, Barmer, Jalor, Sirohi, Pali, Ajmer, Bundi, Bhilwara, Udaipur, Banswara, Chittaurgarh, Kota, Baran, Jhalawar.
	Uttar Pradesh (14/70)	Bulandshahar, Aligarh, Hathras, Mathura, Kannauj, Kanpur nagar, Jalaun, Deoria, Azamgarh, Ballia, Ghazipur, Chandauli, Varanasi, SantRavidas Nagar.
Segment	Assam	Dhubri, Goalpara, Bongaigaon,

International Journal of Data Mining Techniques and Applications
Volume: 05 Issue: 02 December 2016 Page No.159-166
ISSN: 2278-2419

3	(22/27)	Barpeta, Kamrup, Darrang, Marigaon, Nagaon, Sonitpur, Dhemaji, Tinsukia, Dibrugarh, Sibsagar, Jorhat, Golaghat, KarbiAnglong, North Cachar Hills, Cachar, Karimganj, Chirang, Baska, Kamrup Metro.
	Odisha (2/30)	Mayurbhanj, Malkangiri.
	Rajasthan (2/32)	Ganganagar, Churu, Jaipur.
	Uttar Pradesh (7/70)	Saharanpur, Muzaffarnagar, JyotibaPhule Nagar, Meerut, Baghpat, Ghaziabad, Sonbhadra.
	Uttarakhand (13/13)	Uttarkashi, Chamoli, Rudraprayag, TehriGarhwal, Dehradun, Garhwal, Pithoragarh, Bageshwar, Almora, Champawat, Nainital, Udham Singh Nagar, Haridwar.

Total 192 districts of nine high focus states have been divided into three segments. There are 147 districts under segment one, 98 districts under segment two, and 47 districts under segment three. Segment one majorly consists with districts of Bihar, Jharkhand, Odisha, and Uttar Pradesh. All districts of Bihar were grouped under segment one. More than 80% of districts of Jharkhand were also clubbed under segment one. Other than the above, more than half of districts of Odisha, Uttar Pradesh and few districts of Assam, Madhya Pradesh, and Rajasthan have been clustered under segment one. Segment two consists with 98 districts. Maximum districts of Madhya Pradesh, Rajasthan, and Chhattisgarh have gathered under segment two along with few districts from Uttar Pradesh, Odisha, and Jharkhand. Maximum districts assembled under segment three are from Assam and Uttrakhand. All districts of Uttrakhand and around 80% districts of Assam have been categorized under segment three. Segment three is representing districts majorly from hilly regions. Characteristics of segments have been analysed in following subsection to understand each segment well.

3.2 Characteristics of Segments

Average condition of socio-economic and educational parameters and mean percentage of coverage of maternal and child healthcare interventions for all three segments have been studied separately. Comparative differences among segments have been portrayed in figure 2 and 3.

3.2.1 Socio-economic and Educational Characteristics

Figure 2 shows average status of socio-economic and educational parameters for each segment and comparison among the segments. From the figure it can be comprehended that percentage of boys and girls attending school were same among all three segments, around 98%. Percentage of both male and female literacy rate were different. In both parameters, segment three has secured the highest position followed by segment two and then segment one. In case of physical infrastructure condition, except electricity connection rate, all other physical infrastructure parameters have been most in segment three. Maximum electricity connection has been seen in segment two. Other than toilet accessibility rate, condition of other physical infrastructures were worst in districts under segment one. Even in economic condition also, percentage of people owning agricultural land, television, mobile, and motorized vehicle were lowest in segment one. Percentage of people owning land, agricultural land, and motorized vehicles were higher in segment two than in segment three. On the other hand, percentage of people having television and mobile were more in segment three than segment two. It can be concluded from figure 2 that average percentage of socio-economic and educational parameters were poorer in segment one than other segments. Maximum districts out of case study area (147 districts out of 192 districts) belong to segment one.

3.2.2 Comprehensive knowledge of Interventions

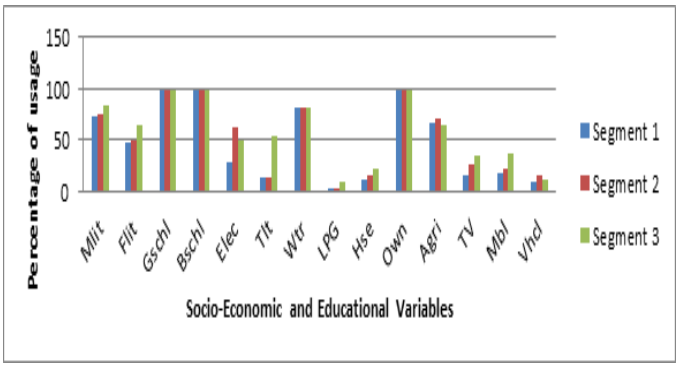

Figure 2: Average percentage of socio-economic and educational parameters for each segment

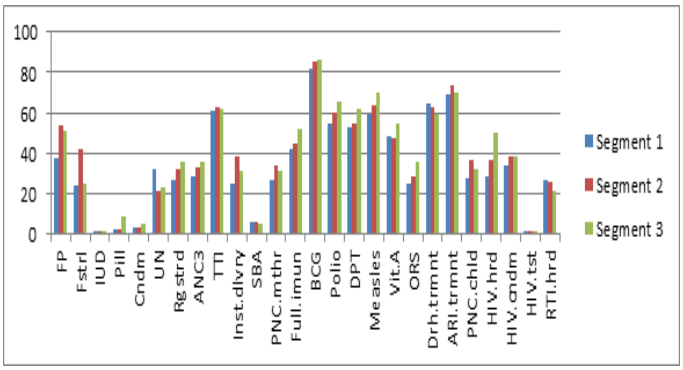

Figure 3: Average percentage of coverage of healthcare interventions for each segment

Figure 3 shows comparison of average values of coverage for all 26 healthcare interventions among three segments. Coverage of all adolescent and reproductive healthcare interventions, analyzed in this study, was worst in segment one than other two segments. Even in case of unmet need also, rate of unmet need was more in segment one than other two segments. Usage of family planning methods and coverage of female sterilization were higher in segment two than in segment three and usage of IUD, pill, and condom were more in segment three than in segment two. Same condition has also been seen in case of maternal healthcare interventions. Other than one intervention i.e. number of skilled birth attendants, all other interventions – mothers registered in first trimester, made three ante natal visits, got at least one TT injection, delivered institutionally, received post-natal care within 48 hours of delivery were lesser in segment one than other two segments.

International Journal of Data Mining Techniques and Applications
Volume: 05 Issue: 02 December 2016 Page No.159-166
ISSN: 2278-2419

Percentage rate of female registered in first trimester and percentage rate of mothers who got three ante natal care visits were lower in segment two than in segment three. Coverage rate of other maternal interventions were higher in segment two than segment three. Coverage of all six interventions, delivered for improving immunity of infants, has been seen maximum in segment three followed by segment two and segment one respectively. Coverage of consumption of ORS was also most in segment three followed by segment two and segment one respectively. Percentage of children who got treatment for diarrhea has been seen highest in segment one and lowest in segment three. Percentage of children got treatment for ARI and got post-natal checkup with 24 hours of birth were maximum in segment two.

Knowledge and activities related to HIV/AIDS have varied from segment to segment. People from segment three have heard about HIV/AIDS were maximum. Maximum percentage of people of segment two knew that consistent use of condom can reduce chance of HIV/AIDS. People went for HIV testing were very low but maximum people were from segment one. Even maximum people of segment one has heard about RTI. Analyzing both the figures, it has been clear that coverage of healthcare interventions at districts under segment one was comparatively worse than other two segments. Separate planning should be done for improvement of healthcare conditions of districts under segment one.

3.2.3 Key Influencing Interventions

For reduction of inequity and improvement of healthcare condition throughout all regions, strategic region-wise intervention designing should be done. Classification and Regression Tree (CART) technique has been used for finding most influential interventions on U5MR and Infant Mortality Ratio (IMR) for all three segments separately. Six pruned regression trees have been developed for indicators and for three segments. These developed regression trees hierarchically represented most relevant interventions through non leaf nodes and at leaf node, average value of mortality rate has been predicted. Figure 4, 5, and 6 are showing regression tree diagrams for all three segments with most important maternal and child healthcare interventions which have best predicted U5MR in those regions. Abbreviated forms have been used in diagrams for utilizing space properly. Details of abbreviated forms have been depicted in table 1.

Figure 4 shows the decision tree diagram developed for presenting the relationship model among U5MR and healthcare interventions for segment one. At districts under segment one, most important interventions were coverage of DPT vaccines, PNC.chld, and Fstrl. It has been comprehended that districts with percentage of consumption of DPT less than 31 percent has observed average U5MR 105.50 per 1000 live births. Districts where consumption of DPT was more than 31 percent and coverage of PNC.chld was less than 21.4 percent have seen average 92.09 U5MR. Districts with DPT more than 31 percent, PNC.chld more than 21.4 percent and Fstrl more than 19.15 percent have seen minimum under-five deaths. Average U5MR seen in those districts were 71.58 per 1000 live births.

Regression tree of most influential interventions on U5MR for districts under segment two is presented in figure 5. Consumption of three doses of DPT vaccines was most relevant intervention for segment two also. All districts under this segment have been classified with DPT. Places with DPT consumption rate less than 41.8 percent have observed average U5MR 97.78 per 1000 live births and with DPT more than 41.8 percent average U5MR was 78.11 per 1000 live births. Figure 6 shows decision tree diagram of most relevant MCH interventions on U5MR for segment three. Here female sterilization was the most influential interventions on U5MR and the split value was 34.4 percent. Districts with Fstrl less than 34.4 percent and more than 51.7 percent of mothers had TT injection (TTI) have seen maximum U5MR (average 78.08 per 1000 live births). Under segment three, districts with Fstrl more than 34.4 percent and ORS consumption by children more than 37.95 percent have seen average U5MR 34.88 per 1000 live births.

Among all MCH interventions DPT consumption by children was the most influential intervention for both segment one and two which have best predicted U5MR. In segment three, best influential intervention was female sterilization. From study it can be understood that not only children and new born interventions but adolescent age intervention also have high impact on children's health. Figure 7, 8, and 9 are showing most relevant interventions for predicting IMR from segment one, two, and three respectively. From figure 7 it can be comprehended that consumption of three doses of DPT was also the most relevant intervention for predicting IMR values of segment one. Other important intervention was consumption of Vitamin A by children. Districts with DPT consumption less than 29.85 percent have seen average 77.11 infant deaths out of 1000 live births which were worst under segment one. Districts with DPT more than 29.85 and vitamin A consumption more than 46.55 percent had seen minimum IMR value under segment one. Average value of IMR of those districts was around 55 per 1000 live births.

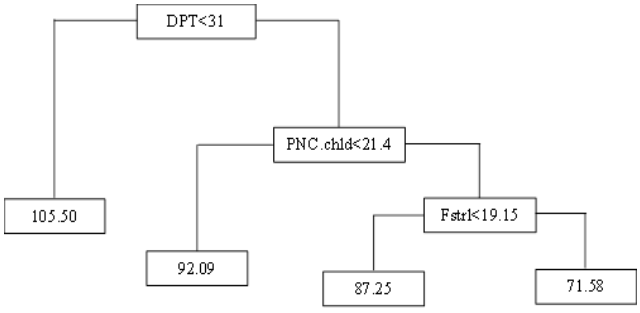

Figure 4: Influential MCH interventions on U5MR for districts under segment one through regression tree diagram

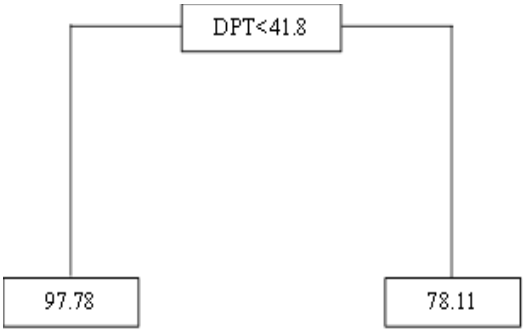

Figure 5: Influential MCH interventions on U5MR for districts under segment two through regression tree diagram

International Journal of Data Mining Techniques and Applications
Volume: 05 Issue: 02 December 2016 Page No.159-166
ISSN: 2278-2419

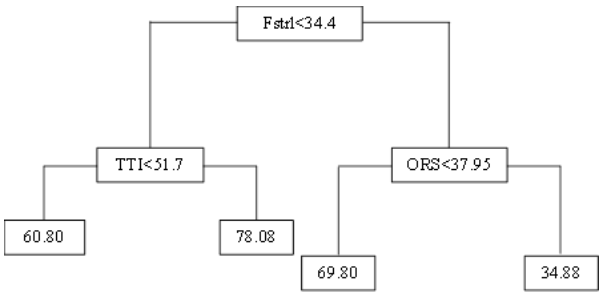

Figure 6: Influential MCH interventions on U5MR for districts under segment three through regression tree diagram

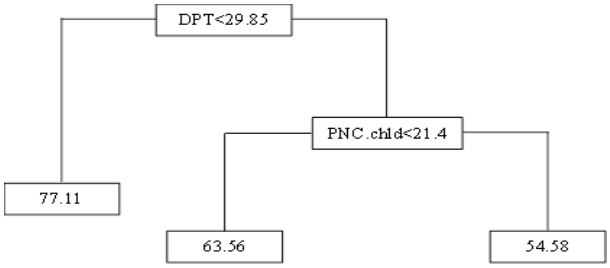

Figure 7: Influential MCH interventions on IMR for districts under segment one through regression tree diagram

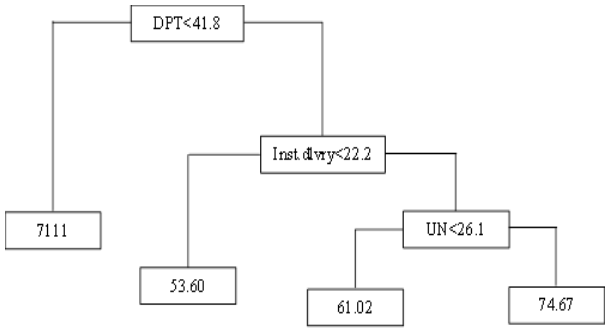

Figure 8: Influential MCH interventions on IMR for districts under segment two through regression tree diagram

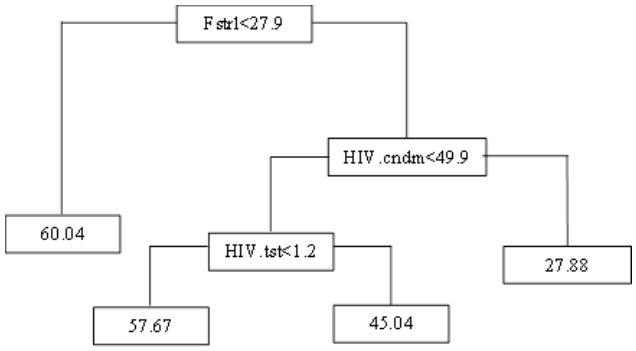

Figure 9: Influential MCH interventions on IMR for districts under segment three through regression tree diagram

According to the regression tree of figure 8, districts with DPT less than 41.8 percent have seen infant death with average around 71 per 1000 live births and districts with more than 41.8 percent children consumed DPT, more than 22.2 percent mothers had institutional delivery, and less than 26.6 percent unmet need have seen least infant deaths under segment two. Figure 9 demonstrates regression tree of most influential interventions of segment three on IMR. In segment three, female sterilization was the most influential intervention. Districts with higher IMR (average 80.04 per 1000 live births) had less percentage of female sterilization. Other interventions which demarcated districts within segment were knowledge of women regarding consistent condom usage (HIV.cndm) and percentage of women ever underwent test for detecting HIV/AIDS (HIV.tst). It has been observed that districts with Fstrl more than 27.6 percent and HIV.cndm more than 49.9 percent have seen average infant death around 28 per 1000 live births. In this segment, districts with HIV.tst more than 1.2 percent have seen average IMR 45.40 and HIV.tst less than 1.2 percent have seen average IMR 57.87 per 1000 live births. For evaluating accuracy level of classifiers, Mean Square Error (MSE) has been calculated for all nine decision tree classifiers. List of all MSEs has been displayed in table 3. Among all, accuracy level of the model which found out influencing MCH interventions on IMR for segment three was best.

Table 3: List of Mean Square Errors (MSEs) for all developed models

Health Indicators	Segment	Mean Square Error (MSE) of the model
U5MR	Seg1	258.79
	Seg2	210.10
	Seg3	136.05
IMR	Seg1	154.91
	Seg2	79.07
	Seg3	62.14

IV. CONCLUSION

For reduction of inequity in healthcare indicators among different regions, in-depth analysis of specific needs should be done and accordingly healthcare delivery should be planned. From this work it has become clear that impact of healthcare interventions on healthcare indicators varies from region to region. In hilly regions, adolescent interventions had more impact on U5MR and IMR than child age interventions.

References

[1] United Nations. The Millennium Development Goals Report 2013. New York, NY; 2013.
[2] J.K. Rajaratnam, J.R. Marcus, A.D. Flaxman, et al. Neonatal, postneonatal, childhood, and under-5 mortality for 187 countries, 1970-2010: a systematic analysis of progress towards Millennium Development Goal 4. Lancet 2010; 375(9730):1988-2008.
[3] M.C. Hogan, K.J. Foreman, M. Naghavi, et al. Maternal mortality for 181 countries, 1980-2008: a systematic analysis of progress towards Millennium Development Goal 5. Lancet 2010; 375(9726):1609-23.
[4] United Nations Children's Fund. Levels & Trends in Child Mortality. Estimates Developed by the UN Inter-Agency Group for Child Mortality Estimation. New York; 2010.
[5] Registrar General of India (RGI), Census Commissioner (CC). Annual Health Survey 2010–11. New Delhi; 2012.
[6] Registrar General of India (RGI), Census Commissioner (CC). Census of India–2011: Provisional Population Totals Series 1. New Delhi; 2011.

International Journal of Data Mining Techniques and Applications
Volume: 05 Issue: 02 December 2016 Page No.159-166
ISSN: 2278-2419

[7] N. Halfon, K. Larson, M. Lu et al. Lifecourse health development: past, present and future. Maternal and child health journal, 2014, 18(2):344-65.
[8] Ministry of Health and Family Welfare G of I. Family Welfare Statistics in India.; 2011. Available at: http://www.2cnpop.net/uploads/1/0/2/1/10215849/mohfw_statistics_2011_revised_31_10_11.pdf.
[9] IIPS. District Level Household and Facility Survey (DLHS-3), 2007-08: India. Mumbai; 2010.
[10] J.R. Quinlan. Induction of decision trees. Mach. Learn. 1986;1(1):81-106.
[11] L. Breiman, J. Friedman, C.J. Stone, et al. Classification and Regression Trees. CRC press; 1984.